

IMPLEMENTASI K-NEAREST NEIGHBOR DENGAN ENSEMBLE FEATURE DAN WORD NORMALIZATION

UNTUK KLASIFIKASI SETIMEN ULASAN E-COMMERCE

Nurnaziah (201510370311139)

Yufis Azhar, S.Kom, M.Kom

Nur Hayatin, S,ST, M.Kom

LATAR BELAKANG

Perkembangan industri *e-commerce* di Indonesia mengalami pertumbuhan yang begitu pesat. Dimana, berdasarkan data *survey* yang dilakukan oleh Badan Koordinasi Penanaman Modal (BKMA), bahwa nilai investor sektor *e-commerce* sebesar 22,6 T. Hal ini menjadikan *e-commerce* sebagai sektor ekonomi yang paling strategis. Salah satunya adalah Tokopedia. Dalam *google play store* sudah tersedia fitur ulasan yang menggambarkan pengalaman pengguna terhadap layanan/produk Tokopedia. fitur ulasan dapat dijadikan tolak ukur yang efisien untuk menemukan informasi. Untuk mengolah informasi tersebut dapat dilakukan dengan analisa sentimen. Tetapi, dalam menganalisa sentimen, terdapat masalah spesifik yang harus diperhatikan. Yaitu ulasan kaya akan fiturnya. Dimana penggunaan fitur *bag of words* saja belum mampu meningkatkan hasil akurasi. Sehingga diperlukan penggabungan fitu (semantik kata dan statistik kata) serta dilakukan normalisasi kata untuk memperoleh hasil akurasi yang maksimal. Ulasan dikelompokan menjadi 2 yaitu; ulasan *positive* dan *negative*, dengan menggunakan metode klasifikasi *k-Nearest Neighbor*.

RUMUSAN MASALAH

1. Bagaimana mengumpulkan data ulasan ?
2. Bagaimana mengklasifikasikan ulasan positif dan ulasan negatif?
3. Bagaimana hasil akurasi dari analisa sentimen ?
4. Bagaimana pengaruh normalisasi kataterhadap hasil klasifikasi?

TUJUAN

Menerapkan metode klasifikasi dan memperoleh hasil akurasi dari analisa sentimen serta pengaruh normalisasi kata terhadap hasil klasifikasi.

ANALISA & PERANCANGAN



KESIMPULAN

Berdasarkan hasil penelitian analisa sentimen terhadap ulasan aplikasi *onlineshop* Tokopedia menggunakan metode klasifikasi *k-Nearest Neighbor* (kNN) dengan kombinasi *Ensemble Feature* dan *Word Normalization*, dapat disimpulkan bahwa penggunaan fitur gabungan (*Ensemble Feature*) memperoleh nilai akurasi yang paling tinggi jika di bandingkan dengan penggunaan fitur secara terpisah. Baik fitur semantik kata maupun fitur statistik kata. Pengujian dilakukan dengan menggunakan nilai *k* terbaik yaitu *k*=45. Sehingga diperoleh, hasil Akurasi menggunakan fitur gabungan statistik kata dan semantik kata (*ensemble feature*) yaitu sebesar (87%), sedangkan akurasi menggunakan fitur semantik kata saja sebesar (65%) dan akurasi menggunakan fitur statistik kata saja sebesar (86%). Selain itu, normalisasi kata (*edit distance*) berpengaruh besar terhadap hasil akurasi dan nilai *k* pada klasifikasi *k-Nearest Neighbor* (kNN). Dimana dari hasil pengujian, akurasi yang diperoleh dengan menggunakan normalisasi kata sebesar 87%, sedangkan akurasi tanpa menggunakan normalisasi kata sebesar 71%.

HASIL PENGUJIAN

1. Pengujian Fitur Semantik

<i>K</i> Terbaik	Akurasi	Presisi Positif	Presisi Negatif	Recall Positif	Recall Negatif	F-Measure Positif	F-Measure Negatif
K-45	0,65	0,62	0,69	0,71	0,60	0,66	0,64
Presentase	65%	62%	69%	71%	60%	66%	64%

2. Pengujian Fitur Statistik

<i>K</i> Terbaik	Akurasi	Presisi Positif	Presisi Negatif	Recall Positif	Recall Negatif	F-Measure Positif	F-Measure Negatif
K-27	0,86	0,87	0,85	0,83	0,88	0,85	0,87
Presentase	86%	87%	85%	83%	88%	85%	87%

3. Pengujian Fitur Gabungan

<i>K</i> Terbaik	Akurasi	Presisi Positif	Presisi Negatif	Recall Positif	Recall Negatif	F-Measure Positif	F-Measure Negatif
K-45	0,87	0,84	0,90	0,90	0,85	0,87	0,87
Presentase	87%	84%	90%	90%	85%	87%	87%

4. Pengujian Normalisasi Kata

<i>K</i> Terbaik	Akurasi	Presisi Positif	Presisi Negatif	Recall Positif	Recall Negatif	F-Measure Positif	F-Measure Negatif
K-45	87%	84%	90%	90%	85%	87%	87%
Presentase	76%	73%	82%	83%	71%	78%	76%